DOCUMENT RESUME

ED 197 387 CS 206 108

AUTHOR Gaies, Stephen J.
TITLE College Freshman Writing Ability in 1963 and 1977: A Pilot Study Comparison.
PUB DATE Nov 80
NOTE 14p.: Paper presented at the Annual Meeting of the Midwest Modern Language Association (22nd, Minneapolis, MN, November 6-8, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Ability: *College Freshmen: *Comparative Analysis: Higher Education: Longitudinal Studies: *Writing (Composition): *Writing Research

ABSTRACT

A pilot study was conducted to compare the quality of essays written by college freshmen in 1963 and in 1977. Twenty-five compositions from each college year were used. The compositions in the two groups of essays were coded according to each writer's sex, high school ranking, and scores on the American College Testing Program. In this way 25 matched pairs of writers were selected and examined. Each of the 50 compositions was evaluated by two of three college English teachers, who rated the overall quality of each composition on a five-point scale. These evaluators were trained in the procedures of the Writing Competency Examination Program, whose eight criteria had been shown to provide a highly workable evaluation format. The data analysis proceeded through a statistical comparison of the two sets of 25 scores (each composition was given a single score, the mean of the two scores it received from its two readers). Although the distribution of scores in the 1963 sample was much narrower, the difference between the means of the two samples was not statistically significant, suggesting that declining writing ability of college students may be something of a myth. Due to the small scope of this pilot study and certain weaknesses in it, such as low inter-rater reliability and the high class rankings of the writers of the compositions, a followup study is being conducted to further test the hypothesis and to explore specific areas in which changes may have occurred. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED197387

COLLEGE FRESHMAN WRITING ABILITY IN 1963 AND 1977: A PILOT STUDY COMPARISON[1]

Stephen J. Gaies
University of Northern Iowa

The study reported on in these pages addresses one of the most publicized and far-reaching issues in contemporary American education: the alleged decline of writing ability among students. The question of whether students today write as well as their predecessors, as well as the parallel concern that large numbers of students leave school without achieving minimum proficiency in written expression, have been widely debated by educators and laypeople alike. The claim has been made that achievement of writing skill, one of the traditional goals of American education and a precondition of a literate society, can no longer be taken for granted. In recent years, the mass media have brought to public attention what many educators claimed they had long recognized: writing disabilities existed on a broad scale not only among high school age students, but also in the college age population. Between 1975 and 1977 in particular, a number of reports in mass circulation magazines, educational journals, and on television asked "why Johnny can't write." The publicity accorded this issue has been so widespread that it has become a piece of conventional wisdom that students today write less and write less well than their predecessors.

Much of the discussion has centered on the causes for this decline, often in conjunction with a search for explanations for declining scores on standardized achievement tests. At the moment we have a large number of possible sources. Some are general, while others relate specifically to language arts instruction in the schools. Of the former, those most frequently advanced include the pervasive influence of the electronic

CS206108

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Stephen J. Gaies
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

media and their implicit devaluation of literacy, the inclusion in testing samples of large numbers of ethnic minority students who in previous years would not have been investigated, and the mood of rebellion against traditional cultural values that was born in the late 1960's. As for the causes specifically rooted in the schools, the abandonment of high school course requirements (in particular, the once-sacred notion that a student ought to take four years of English), overcrowded classes, poor teacher training (which suggests that even those teachers who want to improve their students' writing ability are ill-equipped to do so), the movement away from the teaching of formal grammar and writing mechanics (along with what some claim is an obsession with preserving student "creativity" by not imposing standards of "correctness" on student work): these and others have been repeatedly cited.

There remains considerable debate concerning these "causes" and their relative importance. Far less ambiguous has been the response to the problem. Commissions were formed by federal and state agencies to study the problem. Minimum high school competency requirements in English were adopted in some states. Significantly, writing competency programs were established at numerous institutions of higher learning, both public and private. Even at Harvard, and Expository Writing Program was established, as one writer put it, to provide "remedial writing for the advantaged student" (Dieterich, 1977, 467). At the University of Northern Iowa, only seven years after a two-course composition requirement was abolished, the University Faculty Senate (followed by the Board of Regents) gave final approval to a University Writing Competency Program in early 1977.

In the context of such an immediate and massive response, it has been difficult to consider the issues dispassionately, most of all the question of whether writing ability has in fact declined. Indeed, to suggest that

writing ability has not declined would appear to be an attack on common sense itself. There is, however, an ample number of reasons to do so. First of all, the complaint that "today's" students are in one way or another inferior to those of the past has been a constant lament among educators at all levels. Second, English teachers have traditionally been cast in (or have appointed themselves to) the role of "guardians of the language." From this it could be argued that they apply unrealistic standards in the evaluation of compositions, standards that are tailor-made to claims of widespread writing disabilities. Third, what is meant by "good" writing is hardly a matter of universal agreement. If we know anything, it is that "effectiveness in writing" can mean--and in fact has meant--different things to different people. Finally, and most important, the claim that writing ability has declined rests on a remarkably thin base of empirical data which has been bolstered largely by anecdote.

The study most often cited by proponents of the notion that writing ability has declined is a report of the National Assessment of Educational Progress, an information-gathering project that surveys educational attainment of different age groups in a number of learning areas. This project, which is funded by the National Center for Educational Statistics of the Department of Education, produced a study (Beshoar, 1976) which concluded that in comparison to similar students from 1969, 13- and 17-year-old students in 1974 wrote essays of poorer quality. Two points are of interest here. The first is that what often goes unmentioned is the fact that in this study, the essays written by 9-year-olds in 1974 showed improvement over the efforts of 9-year-olds in the 1969 sample. The second is that from an analysis of the essays, it was concluded that students' mastery of basic writing mechanics (e.g., spelling, capitalization, punctuation, agreement, word usage)--in other words, precisely those things which critics have

commonly claimed would have to be emphasized in any effort to reverse the "downward trend" of writing ability--was not the source of the problem. Thus, the one empirical study of writing trends among school-age writers which has been most widely used in support of the notion that writing ability has declined presents data that is in one respect "self-conflicting" and in another "counterintuitive."

To the data of the NAEP study has been added a mass of essentially anecdotal evidence, often in the form of a claim that students perform less well on essay tests than in the past. Even where these claims are based on data, they present something less than a fully convincing case. Attempts to insure the use of standard tasks and standard scoring procedures (including measurement of inter-scorer reliability) are either poorly made or not made at all. Curiously, those who support the argument that writing ability has declined by way of such anecdotal evidence are often the same people who bemoan the misuse of written language in government and other institutions, thereby suggesting that writing disability is hardly limited to students of today.

In view of all this, it seems quite appropriate to submit the conventional wisdom that writing ability has declined to further empirical verification. In other words, and more specifically, we shall test, under controlled conditions, the hypothesis that the quality of essays written by college freshmen has changed between 1963 and 1977 (in null form, there is no difference in overall quality between essays written by college freshmen in 1963 and those written by a similar group in 1977).

To test the hypothesis, essays written by freshmen at the University of Northern Iowa (UNI) were sampled. In the pilot phase of the study reported on in these pages, twenty-five compositions written in 1963 and twenty-five compositions written in 1977 were used. Thus, in testing the hypothesis, we

asked whether the writers of these compositions can be reasonably supposed to belong to the same population of first-semester freshmen writers or whether first-semester freshmen at UNI in 1977, as judged by the overall quality of their written compositions, should be treated as belonging to a population distinct from that to which the writers of the 1963 essays belong.

The 1963 sample was drawn from a pool of over 1000 compositions written to provide baseline data for incoming freshmen at the State College of Iowa (now UNI) in a study of the effectiveness of college-level instruction in freshman composition (Jewell, Cowley, and Rhum, 1966). The 1977 sample was drawn from 132 compositions written by first-semester freshmen on Form E of an experimental administration of the UNI Writing Competency Examination in August, 1977. Thus, the two samples involved compositions written by incoming freshmen who had had no college-level instruction in composition. Furthermore, the two samples consisted of compositions written under highly similar conditions in response to high similar task formats. In both cases, the topic to be written on demanded a middle level of abstraction, but no specific writing pattern (e.g., comparison, contrast, definition) and no special factual knowledge were required.

Compositions in the two pools were coded according to the sex, high school decile ranking, ACT (American College Testing Program) Composite Standard score, and ACT English Subsection score of the writer. Twenty-five matched pairs of compositions were then selected at random. By "matched pairs" is meant that the writers of the compositions in a pair were of the same sex, had the same ACT English Subsection score, and either the same ACT Composite Standard score or the same high school decile ranking, or both. In eight of the twenty-five pairs, the two writers were matched on all four variables. In the other seventeen pairs, no pair of ACT Composite Standard Score differed

by more than one point, and no pair of high school decile rankings differed by more than one decile. Thus, each pair of writers was matched for three variables; in seven cases the writers were matched on four (see Table 4, p. 8).

The only exception to random selection occurred in the case of compositions whose content in some way clearly revealed whether they were written in 1963 or 1977. Such compositions were to be disqualified. As it turned out, only one composition was so identified.

The compositions selected were retyped without any editorial changes. Each of the fifty compositions was evaluated by two readers (from a pool of three readers), who rated the overall quality of each composition on a 5-point scale (with "1" being the lowest score and "5" the highest). The readers were members of the faculty of the Department of English Language and Literature with previous experience in evaluating overall quality in written composition (although none of the three had been involved in the 1977 experimental administration of the Writing Competency Examination). The readers used the same criteria in evaluating the overall quality of the compositions as they were trained to use in workshop training for the Writing Competency Examination Program (hereafter, WCEP). This procedure was done in part to insure, at least from that point of view, high scorer reliability. More importantly, however, the scoring procedure used has been shown during the first years of the WCEP experience at UNI to be a highly workable evaluation format. A writer, in order to be judged minimally competent in writing ability, must produce an essay which satisfies eight basic criteria: central idea, logical progression, systematic plan of organization, effective transitions, paragraph development, sentence structure, word choice, and mechanics. A composition which satisfies all eight criteria will receive a "5" or a "4"; a composition which is deficient in one or

more of the areas receives a "3", "2", or "1". Readers in the WCEP have available to them a point system by which they can determine whether the criteria of sentence structure, word choice, and mechanics have been met. The readers are, as a result of their workshop training, highly proficient in using this point system and, indeed, in assessing whether or not a composition satisfies all of the eight criteria.

The order in which compositions were rated was rotated so that a composition that was among the first to be read by one reader would be among the last to be read by the other reader. It was decided in advance that in every case in which the two evaluations differed by more than one point, the composition would be read by the third reader; the mean of the closest ratings assigned to the composition would be used in the data analysis.

One final point should be made. The three readers were questioned after they had completed their tasks to see whether they knew what the purpose of the study was. There was no evidence that any of the readers knew that two sets of compositions from different periods of time being compared. They were of course aware that they had rated compositions written for different tasks, but they did not identify any of the compositions as having been written either very recently or many years ago. Part of the success that was achieved in masking the purpose of the study may be attributed to another feature of the evaluation process: the readers were asked to indicate which of the eight criteria a composition failed to meet. This was done, of course, so that the focus of the study on overall quality was to some degree obscured.

The data analysis proceeded through a statistical comparison of the two sets of twenty-five scores (each composition was given a single score, the mean of the two scores it received from the two readers). Score frequencies are presented below in Table 1:

Table 1.

Mean Score	Score Frequency 1963 Sample	1977 Sample
1.0	3	4
1.5	8	10
2.0	4	3
2.5	6	1
3.0	3	3
3.5	1	0
4.0	0	3
4.5	0	1
5.0	0	0

As can be seen from Table 1, the two distributions of mean scores differ in one noticeable way. The distribution of scores in the 1963 sample is much narrower, with 3.5 the highest mean score obtained. The range of scores in the 1977 sample, however, extends to 4.5. Table 2 presents additional information on the performance of the two groups; this information was processed by a Hewlitt Packard 2000 Computer in conjunction with a SPSS (Statistical Package for the Social Sciences) Program to determine whether or not the variances of the two samples could be pooled.

Table 2. Test for Equality of Variances

Sample	Mean	Std. Dev.	Std. Error	F-Value	1-Tail Prob.
1963	2.02	0.699	0.140	2.354	0.018
1977	2.12	1.073	0.215		

The results indicated that the appropriate test for differences in means of the two groups would involve separate variance estimate. This test, which in effect determines the likelihood of the difference in means occurring by chance alone, is summarized below in Table 3:

Table 3. Test for Difference of Means

t-Value	DF	2-Tail Prob.
-0.39	41.27	0.699

Since the difference between the means of the two samples would occur by chance alone almost 70 times out of 100, we conclude that the difference is not statistically significant: in other words, the hypothesis that there is no difference in overall quality between essays written by college freshmen in 1963 and those written by a similar group in 1977 is not rejected.

The pilot study thus suggests that the conventional wisdom that writing ability has declined may be something of a myth. The two groups of writers used in the pilot study were matched on a number of key variables. They wrote essays of roughly similar length in response to highly similar tasks, and their essays were evaluated according to the same set of criteria by highly trained readers who were unaware of the nature of the study in which they were participating. Judging from the performance of these two groups of writers, there seems to be no basis for the claim that college freshmen today write less well than their predecessors.

It would be foolish, of course, to suggest that a study of the scope of the one reported on in these pages is sufficient in the way of empirical investigation of an issue. No such claim is made here. Indeed, this study will be followed by one larger in scope. The follow-up study will involve larger samples, and it will investigate not only the issue of overall quality

in written composition, but also student performance in a number of writing and rhetorical areas: such things as organization, cohesion, word choice, sentence structure, and mechanics. The follow-up study, then, will further test the hypothesis that writing ability has declined; in addition, it will explore specific areas in which change may have occurred.

It is hoped also that certain weaknesses in the pilot study can be minimized or eliminated altogether in further research. Though none of these invalidates the findings of the pilot study, such weaknesses ought to be pointed out.

The most troubling weakness, in the view of the investigator, involved the issue of reliability. Of the fifty compositions that were read by two readers, eight had to be submitted to a third reader. In other words, the two readers differed by more than one point (on a five-point scale) in 16% of the cases. This figure, though lower than what is usually encountered in the grading of themes written for the WCEP (where the figure is generally around 20%), is nonetheless troubling, since it can be so easily be used to support the argument that the evaluation of compositions is inherently unreliable. In the pilot study, all of the elements which generally increase inter-scorer reliability--a clear set of criteria, initial training and extensive experience, and standardization of task--were present; yet certain anomalies in scoring remained. Suffice it to say that the issue of reliability will be closely monitored in the follow-up study which is planned.

The second weakness--although "limitation" might be a better term--of the pilot study is that, with one exception, all of the subjects were ranked in the upper half of their high school class. Thus, the study does not address the issue of writing ability of "poorer" students. All that can be said is that the data base for college freshmen in 1963 was confined to

students ranked in the fifth decile and higher. In view of the difficulty of finding large samples of student writing from fifteen or more years ago which can be easily compared with the written work of students of today, the question of whether writing ability among lower achievers has changed will be difficult to investigate. It is hoped that in the follow-up study, a broader range of student ability can be investigated. Whether or not this can be done depends largely on the data available for study.

A final difficulty resulted from the retyping of the original compositions used in the pilot study. Retyping was deemed necessary, since the paper on which the 1963 compositions were written showed definite signs of age. The problem, however, is that a typed copy presents a very different appearance than a handwritten original. Two of the three readers remarked that errors stand out more on a typed copy; they felt that their evaluations may have been affected by this, although they also added that they were not sure in what way, if any, their evaluations may have changed. As with the previous difficulty mentioned, such a problem may be unavoidable in a comparative study of composition ability such as the one reported on here.

Despite these limitations, the pilot study represents a sound empirical investigation of an issue which, for all the publicity it has generated, remains underexplored. Of the many challenges which face the language arts instructor, the teaching of composition is undoubtedly one of the greatest. It is a myth that students in any era learned to write with universal success. The assessment of the teaching of composition today should be undertaken carefully and dispassionately. One aspect of such an assessment should obviously be a comparison of writing ability today with that of previous years. Such comparisons ought to be based as much as possible on empirical investigation; the alternative is to rely on conventional wisdom, which, as the present study suggests, may not necessarily represent things as they are.

[1] This research was made possible by a Summer Faculty Research Fellowship from the Graduate College, University of Northern Iowa. Additional budgetary expenses were provided by the Graduate College, University of Northern Iowa. The author wishes to thank the Graduate College (in particular, Dr. H. Ray Hoops, former Dean of the Graduate College, Dr. Jan C. Robbins, Acting Dean of the Graduate College, and Ms. Ruth Ratliff, Assistant to the Dean of the Graduate College for Faculty Services) for its generous support.

Beshoar, B.B. "The condition of student writing." American Education, March 1976, 19-22.

Dieterich, Daniel J. "The decline in students' writing skills." College English, January 1977, 466-72.

Jewell, Ross M., Cowley, John, and Gordon Rhum. "The effectiveness of college-level instruction in freshman composition." Interim Report, Project 2188 (Amended), Contract SAE-OE-4-10-053 (Amended), U.S. Department of Health, Education, and Welfare (Office of Education, Bureau of Research). Cedar Falls, Ia.: State College of Iowa (University of Northern Iowa), 1966.

Table 4.

Pair #	1963 Code #	1977 Code #	Sex	1963 ACTEng	1963 ACTCom	1963 HSDec	1977 ACTEng	1977 ACTCom	1977 HSDec
01*	0236	0312	F	26	27	8	26	27	8
02	1049	0052	F	22	23	9	22	24	9
03*	0852	0129	M	19	24	7	19	24	7
04	0461	0079	M	20	20	6	20	20	6
05	0221	0090	F	23	26	7	23	26	8
06	0257	0060	F	26	25	9	26	24	9
07	0186	0136	F	27	27	9	27	28	9
08	0602	0030	M	22	23	6	22	23	5
09	0659	0058	F	25	25	8	25	25	9
10*	0923	0001	F	18	20	6	18	20	6
11	0924	0018	F	25	22	9	25	21	9
12	0683	0089	F	24	22	6	24	23	6
13*	0930	0115	F	32	29	9	32	29	9
14*	0916	0130	F	21	23	6	21	23	6
15	0726	0069	F	20	18	5	20	18	6
16*	0603	0062	M	20	25	5	20	25	5
17*	0981	0048	F	24	22	8	24	22	8
18	0993	0064	F	25	23	9	25	23	8
19*	0422	0066	F	23	24	8	23	24	9
20	0031	0074	M	20	25	6	20	25	6
21	0768	0109	M	22	26	8	22	26	9
22	0308	0111	F	19	17	6	19	17	5
23	0161	0017	F	20	21	9	20	20	9
24	0604	0032	M	17	19	5	17	19	4
25	0516	0021	F	24	26	8	24	26	9

*Pair matched on Sex, ACTEng, ACTCom, and HS Decile Ranking

All pairs are matched on:
1. Sex
2. ACT English Subsection Score
3. ACT Composite Standard Score or High School Decile Ranking or both

No pair of ACT Composite Standard Scores differs by more than one point; no pair of High School Decile Rankings differs by more than one decile.